ז א ב ב ן ־ ח י י ם

הכינוס הבין־לאומי לשיטת המיכון בניתוח הספרותי ובמילונות

ג ד ב ן ־ ע מ י צ ר פ ת י

המיכון של חקר הלשון

תדפיס מתוך „לשוננו" כרך כה
האקדמיה ללשון העברית
ירושלים תשכ"א

זאב בן־חיים

הכינוס הבין־לאומי לשיטת המיכון בניתוח הספרותי ובמילונות

רקע הכינוס ומטרתו[1]

כל מי שנתנסה בעבודה בלשנית או פילולוגית, בהיקף מצומצם או בהיקף נרחב, כגון חיבור מילונים, כתיבת ספרי דקדוק, הוצאת טקסטים על פי כתבי יד או דפוסים רבים, יודע יפה, שרוב הזמן הנדרש לביצוע עבודה מדעית בתחומים האלה החוקר משקיע באיסוף החומר הנחוץ לו כדי לתת לרעיון, שהגה אותו, לבוש וביסוס מדעי. אמנם יש גם באיסוף החומר לצורך העבודה המדעית מן הפעולה האינטלקטואלית, הרוחנית, שהחוקר נקרא לעשותה, והוא עושה אותה מתוך חדוות יצירה, ואינו נכון לוותר עליה, אבל הדברים אמורים אך בעצם מציאתו של החומר ובפענוחו, ואולם מכאן ועד לארגונו, כדי שיהא אפשר להועיל ממנו בעבודה הבלשנית, עדיין הדרך רחוקה מאוד. תחילה עליו לקטלג את החומר, ז״א לרשמו, אם בכרטיסים ואם בצורה אחרת, למיינו ולסדרו. פעולות הרישום, המיון והסידור, שהן מיכניות גרידא, ברבות החומר הן מפרכות את הגוף ומייגעות את הרוח ואף מביאות לידי קיהוי החושים, והן הן האוכלות את רוב הזמן הנדרש לביצוע כל עבודה מדעית בתחומים האמורים, ובוודאי גם בתחומים אחרים. כמה וכמה חוקרים ויוזמי מפעלים מדעיים נתייאשו תוך מלאכתם ומשכו ידיהם ממנה מאימת שפע החומר שראו לנגד עיניהם, ולא סברו – תרתי משמע – כי יהא לאל ידם, ידם ממש, להתגבר עליו.

עבודות בהיקף נרחב, כמו המילון שאנו מדברים בו, אין הבלשן מסוגל כלל לבצע בלא חבר עוזרים, ואם נתמזל מזלו ולרשותו חבר עוזרים, אז בעיה חמורה למדיי מתייצבת לפניו: איך יהא חבר אנשים בחינת איש אחד, חדור אותה ההכרה ואותה הרוח המנחות אותו בכל שלבי פעולתו, ואפילו בשעת הפעולה המיכנית, והן שקובעות את הבחירה שבאיסוף: זה הן, וזה לאו; ובלעדיהן עשוי כל איסוף, ובפרט איסוף של מלים, להיות לקט מקרי לחלוטין ולקוי עד מאוד מצד תועלתו לעבודה שלשמה לוקט, או אפילו חסר כל תועלת שהיא. יתר על כן: מלאכת האיסוף, המתמשכת שנים הרבה בידי אדם אחד בלבד, אף היא נפגמת לעתים קרובות, שכן ברבות הימים המאסף עשוי לשנות את ראייתו בעניין, ואפילו להטות

---

1. בימים 24־26 בנובמבר 1960 נערך בטיבינגן כינוס לשאלות המיכון במחקר הספרותי ובמילונות, וכותב טורים אלה השתתף בו מטעם מפעל המילון ההיסטורי מיסודם של האקדמיה ומוסד ביאליק. הדברים המתפרסמים כאן נאמרו במליאת האקדמיה ביום ט״ז בטבת תשכ״א (4 בינואר 61), והם באים כאן להשלים את המאמר בעניין המילון ההיסטורי שנדפס בלשוננו (כרך כג, עמ׳ 102־123).

מתוך סדר היום של הכינוס:

פרופ׳ ו׳ פוקס (אכן): על ניתוח מתימטי של הלשון.

פרופ׳ א׳ ולטר (דארמשטאדט): מכונות חישוב אלקטרוניות והשיבותן למדע.

פרופ׳ ר׳ בוזה (גאלארטה): כלים אוטומטיים חדשים לחקר הלשון וללקסיקוגראפיה.

פרופ׳ ו׳ שאדיוואלד (טיבינגן): הערכה של שיטות המיכון בלקסיקוגרפיה.

את מגמת מחקרו בכלל; נמצא, שהחומר שכונס בזמנים שונים שוב אינו אחיד הומוגיני, ועובדה זו פוגמת בסופו של הדבר גם בעצם תוצאת המחקר. לפילולוגים אולי כדאי להזכיר מקרה, שאינו משמח, אך מצוי למדיי. כשפילולוג מכנס גרסאות מכתבי יד רבים, ומספרן עולה לאלפים, הוא לעתים מוכרח להחליט, שיסתפק במבחר גרסאות לפי הערכת חשיבותן או חשיבות כתבי היד שלפניו; אך מה עולה לו בסוף עבודתו, כשבהערכתו את כתבי היד חל שינוי, והלא בהערכתו הראשונה תלוי מבחר הגרסאות שלוקטו בעמל רב. דברים אלו מפורסמים הם, ונתנסו בחומרתם בייחוד אנשים, שהוטל עליהם לבצע עבודה בהיקף נרחב בצוותא. הרבה תקנות תקנות נתקנו למנוע תוצאה כזאת, בעיקר במפעלים לקסיקוגראפיים גדולים, אבל אין לומר, שהתקנות הייתה להן הצלחה גמורה: ככל שמרובים העובדים, כן יקשה למנוע ריבוי ראיות בחומר, אפילו בהדרכה מרכזית קפדנית, וככל שמתמשך האיסוף, ודאי שגדלות האפשרויות של מידות שונות בראייה ובהערכה, וכך נשארה השאיפה לאובייקטיביות הדוקומנטציה עדיין שאיפה שלא הושגה עם כל התקנות שהותקנו.

והנה לפני שנים אחדות עלו באופק הבלשנות כלים חדשים, העשויים לגאול את הבלשן ואת הפילולוג מן העבודה המיכנית המפרכת ואף לפתור את שאלת אובייקטיביות הדוקמנטציה ושל קצב השגתה פתרון גמור. עלו באופק כלים, שישיירו לבלשן אך את העבודה הרוחנית היוצרת. כוונתי למכונות מיון וחישוב חשמליות ואלקטרוניות, שהומצאו כלפני חמש עשרה שנה והן הולכות ומשתכללות בקצב מהיר ביותר. מכונות אלו, שנוצרו במחשבה תחילה לשמש בתפקידים אמרכליים, הגו אנשים אחדים בארצות שונות את הרעיון לרתום אותן לעבודה הספרותית והבלשנית. מאלה אזכיר אחד, שפעולתו ידועה לי יותר: האב בוזה, פרופיסור לפילוסופיה בגאלאראטה שעל יד מילאנו, קיבל על עצמו בשנת 1946 לחקור את משמעותה ושימושה של מלת in הלאטינית בכתבי תומאס מאקווינו. הוא החל לשלות מלה זו מים המלים שכתב אותו סופר – בערך 13 מיליונים. משרשם עשרות אלפי מראי מקומות, ראה שעדיין רוב הדרך לפניו. העלה אפוא על דעתו, שמא תעשה המכונה עבודה זו בשבילו. למן הרעיון ועד להקמת המרכז הראשון בעולם ל״ניתוח ספרותי בשיטת המכונות״ בשנת 1957 עברו אפוא 11 שנים של גישושים וניסיונות. כפי שאנו רואים, דווקא הצורך הלקסיקוגראפי הוא שהוליד את הרעיון הזה במוחו של אותו החוקר. במקומות אחרים צרכים אחרים מתחום הבלשנות: ניתוח תחבירי, ניתוח סגנונו של סופר, ועל כולם תרגום מלשון ללשון – וזה צורך מעשי מאוד, שכנגדו הצורך בידיעת השימוש של in בלשונו של תומאס מאקווינו אינו נחשב כלום, לפחות בעיני מייצרי המכונות – אלה דחפו את הדחיפה לפתח את שיטת המיכון של העבודה הבלשנית. משנת 1957 לבשו כל הגישושים והניסיונות צורה מוחשית מאוד, והיום, במוצאי שנת 1960, כבר יש מרכזים אחדים בעולם, המפתחים שיטה זו בניתוח הספרותי ובעבודה הבלשנית. בינתיים הולכת שיטה זו ותופסת מקומה בתחומי מדע אחרים, בפרט במדעי הטבע ואף ברפואה.

כיוון שהיום יש כבר כאמור כמה וכמה מרכזים לעבודה בלשנית בשיטת המכונות, וכבר נפרדו אלה מאלה בכמה דרכי עבודה ובסוגי המכונות, יזם פרופ׳ בוזה את הכינוס, שמטרותיו המפורשות היו אלה:

א) בירור הדרכים לסילוק כמה פגמים והגבלות בשיטה האמורה, המותנים במבנה המכונות, כגון: מיעוט האותיות ושאר סימני הכתב במערכת המכונה הכותבת;

קשיים במעבר מכתב אחד לזולתו, כגון מכתב לאטיני לכתב עברי וכיו״ב;

ב) סטאנדארדיזאציה של שיטות, מידות וקודים ומונחים כדי להקל את המשא והמתן הבין־לאומי בתחום זה;

ג) הפחתת ההוצאות הכרוכות בביצוע העבודות בשיטת המכונות, כי בשעה זו הן מרובות מאוד, ומונעות מספר גדול של בלשנים מלהשתמש בשיטה החדשה בביצוע עבודותיהם.

אלה, כאמור, היו המטרות המפורשות, והן בעצם טכניות, אבל נדמה לי, שלא אטעה באמרי, שהמטרה הכלולה – ושמא היא עיקר בעיני המזמינים – הייתה לעורר בפילולוגים ובבלשנים את העניין בשיטה החדשה ולקנות את לבם, ובייחוד את לב מנהלי מפעלים בלשניים, לשיטה החדשה. כי עזרתם של הבלשנים נחוצה למפתחי המכונות האמורות, ואולי אף ליצירת מכונות מיוחדות מותאמות לעבודה הפילולוגית והבלשנית. מידת היעילות שבשיטה החדשה תלויה הרבה מאוד בצורה, שהחומר הלשוני מוגש למכונה. אם הבלשן רוצה להפיק את מלוא התועלת מן המכונה, הוא נדרש לניתוח מדוקדק יותר של מבנה הלשון, של כל לשון לעצמה, ולסיסטמאטיזאציה של הגורמים הלשוניים. כי המכונה אומרת לבלשן: קח לך – משלך. אף זאת: המכונה עצמה העלתה שאלות חדשות התובעות מן הבלשן תשובה.

באי הכינוס ודרך עבודתו

האנשים שזומנו לכינוס הזה שלושה סוגים הם:

א) מלומדים שכבר נתנסו בשיטת המכונות בעבודה הבלשנית, בעיקר לקסיקוגראפית;

ב) מלומדים שנתנסו בעבודה הלקסיקוגראפית בשיטה המקובלת, אבל לבם פתוח ללמוד מה שהשיטה החדשה מציעה להם;

ג) מומחים היכולים להסביר באלו תנאים עשויה שיטת המכונות להועיל לעבודה הפילולוגית והבלשנית.

השתתפו כ־50 אנשים מ־7 ארצות: איטליה, אנגליה, ארצות־הברית, בנלוקס, גרמניה, ישראל, צרפת, ונציג של לא־מולדת – כמדומה לי בלגי – ראש EURATOM של או״ם. נציג אונסקו ונציגים של כמה ארצות אחרות נתעכבו מלבוא. המקום גרם, שרוב המשתתפים היו גרמנים מאוניברסיטאות שונות. המשתתפים היו באי כוח של אקדמיות, של חברות למדעים, של אוניברסיטאות ושל מכונים שונים. אזכיר רק אחדים שיש לנו עניין מיוחד להכיר עבודתם, והם עורך Thesaurus Linguae Latinae; עורך Thesaurus Graecus; עורך Mittellateinisches Wörterbuch; עורך Goethe Wörterbuch; ראש המכון Zentralstelle für Terminologieforschung und; בא כוח של Vetus Latina praktische Sprachfragen, Bonn; מנהל עבודה לקסיקוגראפית בגרמנית העתיקה מקמברידג׳; עורך Woordenboek der Nederlandsche Taal, שהכנתו ופרסומו נמשכים זה מאה שנה; ראש המרכז באיטליה; ראש המרכז ללקסיקולוגיה בביזנסון שבצרפת. מרכז זה עושה את העבודה בשיטת המכונות בשביל כל אוניברסיטאות צרפת.

הכינוס שנמשך 3 ימים מלאים וגדושים נפתח, שלא כפי התכנית, בפגישה עם עיתונאים – (press conference). ובה נתבקשו אנשים אחדים לספר לעיתונאים על מוסדותיהם או מפעליהם ועל מטרת בואם לכינוס. לאחר דברי פתיחה של

היושב ראש קראתי מן הכתב דברים על הלשון העברית ועל המילון, שהאקדמיה ומוסד ביאליק החלו בהכנתו. אחריי דיברו נציג איטליה ונציג בנלוקס, עורך Vetus Latina ומתימטיקאי. בכינוס עצמו היו 4 הרצאות: שתיים של מתימטיקאים, של ראש המרכז בגאלאראטה ושל עורך המילון של גיטה (אגב אורחא, הוא בעל קתידרה לפילולוגיה קלאסית בטיבינגן). ההרצאות של המתימטיקאים נתלוו הדגמת העבודה במכונות, ולא היה אחריהן ויכוח, כי לא היו מי שיעמדו עמם בוויכוח, אבל היו שאלות והובעו משאלות. מובן מאליו, שנשאלו שאלות העברית: ראשית כול, כיצד להורות למכונות שיבחינו בטקסט העברי בין הומוגראפים, כלומר מלים הכתובות בשווה כגון כָּתַב, כְּתָב, כַּתָּב וכיו״ב. בכל לשון יש תופעה זו, אבל בעברית ובכמה לשונות שמיות אחרות זו רעה חולה, שוועדת הכתיב מבקשת לה תרופה, ואפשר שהמיכון ידרוש תרופה זו בהקדם. שנית, איך אפשר להביע את סימני הניקוד במכונה, אלה דאגות, שבעלי המילונים האחרים היו פטורים מהן (ערבים ומחברי מילונים לערבית לא היו בכינוס). המומחים ציפו לשאלות כאלה וביקשו לשמען, וויכוח ממשי מעניין מאוד התפתח לאחר הרצאת עורך המילון של גיטה.

אם מבין אני לסכם במלים מעטות, מה שעלה מהרצאותיהם של המתימטיקאים — הרצאותיהם יתפרסמו בפרוטוקול וישמשו חומר לעיון — הרי הוא זה: א) כבר נעשו ניסיונות להשתמש במכונות לקביעת אוטנטיות של מקור ספרותי, אם יש לכך כמה נקודות אחיזה. העניין עדיין בשלב ההתחלה, אך אזכיר מאמר שנכתב בשנת 1958, והוא עוסק בניתוח סגנונם של כתבי אפלטון לשם בדיקת האוטנטיות או הכרונולוגיה שלהם[2]; ב) בזמן הקרוב עשויים להסתיים הניסיונות לייצר מכונה, שתהיה מסוגלת לקרוא בעצמה טקסט, הכתוב באותיות דפוס — ואפילו מנסים את הדבר לגבי כתבי יד — וכשהמכונה תקלוט את הטקסט, היא עצמה — ללא עזרת אדם — תכתוב את הפתקאות ותגישן ערוכות לפי מה שיידרש לבלשן; דבר זה יהיה מהפכה בעבודת המוציאים טקסטים, שכן עם מסירת כל כתבי היד למכונה, זו תקלוט את כולם ותכתוב את הטקסט אך פעם אחת ותמציא את כל חילופי הנוסחאות עם הטקסט. חושבני, שדי לי להזכיר אך שני דברים אלה, שאם יתגשמו, לא יהיו בפינו מלים להעריך את חשיבות העזרה, שתוגש בהן לעבודה הפילולוגית והבלשנית. הצד השווה לכל סוגי העבודה בשיטת המכונות הוא — השגת שלמות הדוקומנטציה בזמן קצר ביותר וחישוב התוצאות בשיטות מתימטיות־סטאטיסטיות. הכמות העצומה של הדוקומנטציה, שאך מכונה מסוגלת להתגבר עליה, מתגבשת ע״י תהליכי חישוב סטאטיסטיים לאיכות גבוהה ואובייקטיבית יותר.

מסקירת ראש המרכז באיטליה אציין, שבאותו המרכז הוכנו רשימות מלים וקונקורדנציות מתוך טקסטים לאטיניים, עבריים (מגילות ים המלח) ומן הכרך השלישי של Farbenlehre של גיטה, וכן — וזה נראה לי דבר חשוב — נעשה ניתוח פונולוגי של הלשון האיטלקית שבימינו. המחקר נעשה כתיזה לחיבור הדוקטור

---

2. השווה A.D. Booth, L. Brandwood, J.P. Cleave: Mechanical Resolution of Linguistic Problems, 1958; Stylistic Analysis pp. 50-65. — שם, עמ׳ 54 נאמר: The style in which a person writes is a personal characteristic, and he can be identified by it just as surely as by his fingerprints, provided that he does not deliberately set out to disguise it.

בפאדובה, המורה ותלמידו[3] שניהם השתתפו בכינוס. כן נעשה ניתוח שלאותה הלשון בפי ילדים עד גיל 6 בשיטה זו, אף הוא באוניברסיטת פאדובה. כל מי שהתמודד ברישום מדויק של דיאלקט כלשהו, יוכל לצייר לעצמו לאיזה דיוק עשוי רישום להגיע, כשתשוכלל האפשרות להעביר את הטקסט ישר מן הסרט של הרשַמקול אל המכונה לשם ניתוח, ומה ייחסך לו ביגיעה. אולם בשבילי, ודומה אני בשביל רוב הבלשנים, מרכז העניין הייתה הרצאת העורך של מילון לגיטה. בעבודה הוחל בשנת 1947, והיא נעשית מטעם האקדמיה בברלין. עסוקים בה 15 אנשים בשלושה מרכזים: טיבינגן, ברלין ולייפציג. לפני זמן מועט נסתיים האיסוף. נוצלו כל כתבי גיטה, שבמהדורת ויימאר הם כלולים ב־142 כרכים. מספר הכרטיסים שלושה מיליונים. אמנם האיסוף נסתיים, אך לשם ביקורת הועבר כרך אחד מכתביו דרך המכונות. חבל, שהכינוס הוקדם בכמה חדשים, ולא ניתנה שהות למפעל הזה להשוות את תוצאת עבודתה של המכונה אל תוצאת עבודת הידיים באותו הכרך. נימת העורך הייתה קצת ספקנית לגבי התועלת שבמכונה, והוא שאל, התוכל המכונה להגיש לעורך המילון חומר כה מחושב ומעובד, ששוקעה בו מחשבת אנשים מצוינים במשך שנים? אודה על האמת, לא יכולתי להשתחרר מרגש הקנאה למראה הכרטיסים המעובדים, ולמראה הערכים שהוכנו לדוגמה ומהם אף פורסם הערך gesund בחוברת לעצמה. חסידי השיטה במכונות חישבו שכרן והפסדן של שתי הדרכים, המקובלת והחדשה, זה כנגד זה, גם מן הצד הכלכלי. לי ברור, והוזכר הדבר בוויכוח, שאין שתי הדרכים סותרות זו את זו, אלא משלימות זו את זו. אין המכונה יכולה לתת את העבודה הרוחנית המשוקעת בכרטסת ענק כזו, אבל היא יכולה לעשות את העבודה המיכנית בזמן פחות הרבה, ואמנם אופייני הוא, שבדברי התשובה נסוג המרצה והודה בחשיבות הדרך החדשה לעבודה הלקסיקוגראפית, ואף נימק הודעתו בעובדה, שנתן הוראה להכין כרך אחד בשיטה החדשה.

אולי יותר ממה שלמדתי מן הכינוס עצמו למדתי משיחות עם באי הכינוס, שסיפרו על בעיותיהם בעבודתם, וראיתי שאמנם סדנא דארעא חד הוא. כדאי אולי לציין, שעניין העברית הוא מטבע, ששערו חריף בעולם, אפילו בעולמם של אנשי המדע והלשון. באתי אל הכינוס כנציג של מפעל, שאך בא לעולם, ואני מקווה שיהיה בן־קיים, וראיתי להביא עמי קצת אינפורמציה על המפעל. התקנתי את תכנית המפעל שאושרה באקדמיה לטעמו ולהבנתו של מי שאינו בקי בעברית, וזו תורגמה לאנגלית כחוברת בפני עצמם. בדברי ההקדמה לחוברת סופר על האקדמיה ועל מוסד ביאליק והוזכר המילון המקראי. חוברת זו היתה האינפורמציה היחידה, שהגיש מפעל מילוני — ThLL, ThG והמילון לגיטה לא היו זקוקים לכך — והיא תידפס בפרוטוקול של הכינוס. עורך ThG העיר, שיש לו בעיות דומות מאוד לאלה שיש לנו. מפעל זה קיים שנים מספר, ועדיין לא החל לכתוב את כרטיסי ההבאות, עדיין הם שקועים בהכנות. להפתעתי היו אחדים, בעיקר מן הכמרים והתיאולוגים, שידעו מקצת עברית, ובעיות של המילון העברי לא היו זרות להם לגמרי.

המסקנות וההחלטות

הכינוס נסתיים בקבלת החלטות, שאזכירן מיד, אע״פ שאיני בטוח, אם דקדקתי די הצורך במסירתן. עדיין אין הפרוטוקול של הכינוס לפניי.

3. והם פרופסור Tagliavini ודוקטור Zampolli.

א) השם ״קולוקביום לשיטות העבודה במכונות בספרות ובלקסיקוגרפיה״ שאליו הוזמננו, שונה לשם ״כינוס בין־לאומי וכו׳״.

ב) בעוד 3 שנים יקוים כינוס שני לסקור את התפתחות העניינים, שכן הכול היו סבורים, שתוך שלוש שנים תחול התפתחות גדולה בענף זה, ויהא צורך להכיר את השינויים, שיחולו במבנה המכונות ובסוגיהן. בכינוס הבא יוחלט, אם יש מקום לקיים מוסד נפרד, ״כינוס של העובדים בשיטת המיכון״, או שהוא יסונף כסקציה לאחד המוסדות הבין־לאומיים, כגון ״כינוס הבלשנים״ או כינוס העוסקים ב״תורת האינפורמציה״ וכיו״ב.

ג) יפורסם דין וחשבון על הכינוס בפרסום נפרד. על צורת הפרסום יחליט יו״ר הכינוס עם בא כוח הפקולטה למדעי הרוח של טיבינגן.

ד) הכינוס ינסח דרישה אל מייצרי המכונות, שיעמידון בתנאים נוחים לעבודות הבלשניות ויסב את תשומת לבם לפגמים הצריכים תיקון, כדי שיהיו המכונות מתאימות יותר לעבודה הבלשנית.

אלה ההחלטות הפורמליות, והמסקנות, שהסקתי מן הכינוס, ושותפים עמי לפחות אותם בעלי המילונים, ששוחחתי עמהם באורח פרטי, ואותם שגילו דעתם בוויכוח, הן אלה:

1) להבא לא תיתכן עבודה בלשנית, בפרט לקסיקולוגית בהיקף נרחב, שלא בשימוש המכונות. יהא זה מסוכן מדאי לסמוך על דוקומנטציה נבחרת במקום דוקומנטציה שלמה. במאמר שקראתי בינתיים אני מוצא משפט מסכם זה[4]: it is also true that computers make it possible to submit all the elements of spoken and written discourse to a systematic study impossible with the individual, manual methods of yesterday.

2) מסתבר שהשימוש במכונות ניקוב ייסוג מפני המכונות האלקטרוניות, המצטיינות במה שהן מסוגלות להחסין בקרבן כמות עצומה של חומר לשוני בגלילים קטנים, בעוד שמכונות ניקוב הפועלות ע״י ניקוב כרטיסים מחייבות קיום כרטסות עצומות התופסות מקום רב. מכונות אלה מצוידות לא רק בזיכרון עצום (מלאי החומר שנאצר בהן), אלא שע״י יכולת קומבינאטורית הן מבצעות פעולות חשיבה מסוימות; ועליהן יאה המאמר שיוחס לתיאודור הויס: Das sind Maschinen die zwar denken können, aber keine Phantasie haben.

3) מפעל בלשני צריך שיהא מעובדיו איש, שהוכשר במיוחד להכין את החומר למכונה, ולזה נחוץ כישרון טכני מסוים וידיעות מסוימות במתימטיקה וסטאטיסטיקה. האיש חייב, כמובן, להיות קודם כל בעל הכשרה בלשנית. אזכיר, שעורך המילון ההולנדי ישב חדשים מספר בביזאנסון כדי ללמוד דרך העבודה באותו המכון.

4) השאלה, כיצד תבחין המכונה את ההומוגראפים – ובעברית זו בעיה מיוחדת במינה בגלל היעדר תנועות בכתיב – לא נפתרה עד עכשיו, והיא מחייבת עדיין התערבות האדם; על נקודה זו ימשיכו בוני המכונות ליתן את דעתם. ומכאן נובעת

---

4. מובא ע״י Th. A. Sebeok בחוברת שהוגשה לכינוס: Notes on the Digital Calculator as a Tool for Analyzing Literary Information (Center for Advanced Study the in Behavioral Sciences and Indiana University).

5) השאלה, שאף היא צריכה עיון: באיזה שלב של העבודה יש להשתמש במכונות. שתי אפשרויות הן:

א) פְּרֵיאֶדִיצְיָה, כלומר התקנת הטקסט התקנה בלשנית קודם למסירתו למכונה ומתן הוראות מדויקות למכונה, ואז יושג חומר ממוין יפה יפה, כשם שנעשה הדבר ב־Prologomena של קאנט. בדרך זו ניתן לכל מלה של הטקסט סמל מיוחד, וזו עבודת הכנה עצומה, או

ב) מסירת הטקסט למכונה כמות שהוא והשקעת העבודה הרוחנית בחומר הגולמי, שהמכונה תשפיע לבלשן.

אשר לעברית – בראייה ראשונה הדרך הראשונה, ז״א פריאדיציה, נראית לי, שאם לא כן, יוצף כותב הערך או העורך בחומר גולמי של הומוגראפים עד לאין שיעור. אבל – כאמור – עניין זה עוד צריך לימוד, וחושש אני, שעד שלא נעביר אחד הטקסטים דרך המכונה, לא נוכל להכריע מה טוב לנו. ומכאן אל

לקח הכינוס לעבודתנו

מן הראוי להעלות פעם אחת פרשת העבודה במילון לפני המליאה, לסקור את שעשינו, ועוד יותר את שנמנע ממנו לעשות, וזו פרשה לעצמה הצריכה דיון לגופה. אע״פ כן אי אפשר לי, שלא אציין כמה עובדות טכניות גם בהקשר זה.

תכנית עבודתנו נתקבלה בינואר 1959. על פיה היינו צריכים להתחיל בעבודות הכנה משני כיוונים, כשלרשות המילון 3 עובדים מדעיים העוסקים במפעל כל אחד יום תמים. על פי ההערכה, שאמנם לא נתחייבתי עליה, הייתה עבודת ההכנה צריכה להימשך 3 שנים עד שעת כתיבת כרטיסי ההבאות מן המקורות באפריל של אותה שנה נכנסו בשירות המפעל שני עובדים, האחד בשני ימי עבודה, האחר בשלושה ימי עבודה – יחד 5 ימי עבודה. כשלושה חדשים היו דרושים, כדי שהעובדים ילמדו את בעיות המילון, בעיקר את תפקידיהם המוגדרים לתקופת ההכנה. משנכנסה האקדמיה לבניינה, נוצר מקום העבודה, וצויד בספרים, ובמקצת גם ברהיטים הדרושים, ועדיין לא הושלם הדבר. בספטמבר של אותה השנה יכול אחד העובדים להוסיף 2 ימי עבודה נוספים, ומאז ועד נובמבר נתנו שני העובדים יחד 7 ימים בשבוע. בסוף נובמבר נצטרף עובד שלישי – 3 ימי עבודה – ומצב זה נמשך עד יוני 1960. אלא שמפברואר 1960 החלו הפרעות בעבודה, ובסופו של דבר נאלצנו להפסיק את עבודתנו ביולי, ורק עובד אחד, שלגביו הייתה התחייבות מפורשת מצד המזמין, מוסיף לשרת את המפעל ביום עבודה תמים. כך המצב עד עכשיו, ואיני בא בהאשמות אישיות, אלא בציון עובדות. דברים אלה בוודאי אומרים לחברים, היכן אנו בעבודתנו, ומה עבר על המפעל עם היווסדו. מתוך מפעלים לקסיקוקרפיים הרבים, הנעשים בגרמניה ללשונם וללשונות הלאטינית והיוונית ואחרות, אזכיר נתונים אלה של המילון לגיטה, שהועלו בכינוס. הוחל בו ב־1947. עסוקים בו 15 אנשים קבועים, נוסף על סטודנטים מסייעים באופן ארעי, בשלוש ערים. בשנת 1960 נסתיימה הכנת הכרטסת. היא עלתה שלושה מיליונים מארקים, ז״א מיליון וחצי מיליון לירות, לפי השער הרשמי. הווי אומר: מדי שנה בשנה הוצאו כמאה ועשרים אלף לירות להכנת הכרטסת בשביל מילון של גיטה. אינני יודע, בכמה עולה להם הכנת המילון הכללי מיסודו של גרים. אם רק נעמיד את הסכום העלוב של 24 אלף לירות לשנה (שלא נוצל בשנה הראשונה בשלמותו) בשביל שני המילונים (המקראי

והכולל), ונמוד את העבודה, שצריכים אנו להשקיע בהכנות, עד שיהא אפשר לכתוב את הכרטיסים מן המקורות, בעוד שהעבודה במילון של גיטה נעשית ע״פ מהדורה מוכנה – יבין המבין, מה מגוחך הסכום שהועמד לעבודת המילון. בסכום כזה בשום פנים אין אפשרות ליצור צוות של עובדים, שיהיו נתונים בכל כוחם ולבם למפעל; אין להקים מערכת, שתיקח על עצמה תפקידים מעשיים והם מרובים. בקיצור אין לדבר על מפעל המילון כראוי לאקדמיה ללשון. ואם בכל זאת אומר מלים אחדות המכוונות לעתיד עבודתנו, הרי רק מתוך אמונה, שהמצב ישתנה מיסודו שינוי גמור. אם לאו – כל זה יישאר בגדר מלים, יפות או רעות, שלא כדאי לבזבז עליהן זמן בוויכוח.

מטבע הדברים הוא, שכל זמן שהותי בכינוס ולאחר מכן הרהרתי בתכנית עבודתנו, אם כשרה היא גם לתנאי השיטה החדשה, ובייחוד בדקתי את שאלת הכרטסת, שעליה נאמר בתכנית שלפני 3 שנים [5] כי היא טעונה בדיקה לאור השיטה החדשה, וזו הייתה ההזדמנות וגם המטרה שלשמה יצאתי להשתתף בכינוס שנתקיים בגרמניה.

שלושה שלבים יש להבחין בהכנת המילון: א) הכנת המקורות למילון וניצול (dépouillement) של הספרות המקצועית, הדנה בפירושי המלים ובטקסטים שיתייסד עליהם המילון. זו מפוזרת בעיקר בעשרות כתבי עת; ב) ניצול המקורות עצמם; ג) כתיבת המילון. השימוש בשיטה החדשה מכוון לשלב השני, ואפשרי אף בו. אבל הוא נוגע גם בדרך העבודה בשלב הראשון, שבו אנו עסוקים. השלב הזה צריך שייעשה בהרחבה ביתר כוחות עבודה, וצריך להימשך זמן ארוך יותר משהובא בחשבון. אע״פ כן מבחינת הזמן הדרוש להכנת המילון אין הארכתו של השלב הראשון נחשבת כלום, שכן השלב השני, עשוי להימשך בשיטה הרגילה לפחות 20 שנה, ייעשה בקצב מסחרר ממש בשיטה החדשה. אלא ששיטה זו מחייבת הוצאות גדולות. בשלב הראשון אנו עסוקים בקביעת מבחר המקורות ובבירור בעיות ספרותיות הכרוכות בהם. שאלת המבחר היא לעולם קשה ומחייבת דקדוק רב, שלא יהא מקרי. ואולם אם נתכוון לשיטה החדשה, נוכל להיות ליבראליים יותר במבחר. אם לא ננצל את כל המקורות, שישנם בתחום תקופה מסוימת, נוכל בכ״ז להקיף הרבה יותר ולהגיע אל מקסימום החומר בזמן קצר יותר. זה מקל הרבה מן האחריות שיש במבחר, אבל מרחיב את בירור הבעיות הספרותיות של מקורות רבים יותר. שנית, וזה בשבילי העיקר, כאמור, שיטת הפריאדיציה נראית לי יותר, פירושו של הדבר, שעבודת ההכנה עד לניצול לא תסתפק בבחירת המקורות ובבירור בעית הכרונולוגיה שלהם וכיו״ב, אלא תהא חייבת בהתקנת הטקסט עצמו בצורה שהמכונה תחזיר אותו ממוין מקסימום המיון שהיא מסוגלת לו. משמע, שהכנה בשיטה זו היא חלק מן העריכה. יהא צורך לנקד מלים בהבאות לסמן הומוגראפים, לציין בקודים חלקי דיבר שונים של כל טקסט שישלח למכונה. מכאן שיחס העבודה המשוקעת שבין הכנת החומר ובין כתיבת הערך עצמו משתנה בדרך החדשה לעומת מה שחשבנו.

המפעל חייב לדאוג, שיוכשר איש מבין עובדיו שישנם או שיהיו, להתקין חומר לשוני למכונה. הוא צריך להיות איש הביניים, בין הבלשן ובין מתקין התכנית של

5. לשוננו כג, עמ׳ 122.

המכונה, שכן עדיין אין מלמדים במחלקות לבלשנות את התקנת התכנית (דרך אגב אזכיר כי לימוד זה הוכנס בכמה מוסדות השכלה טכנולוגיים, והוא חובה על כל תלמיד). יש מבחן מיוחד הקובע, מהן התכונות הדרושות מאיש כזה, שיצליח בעבודתו. בשעת התקנת הטקסט למכונה בצריך יפה לדעת, איך ניתנת המכונה לניצול יעיל.

ובכן מסקנתי היא: דרך העבודה, שקבענו בתכנית ושאנו הולכים בה, עומדת בהחלט גם בשיטה החדשה. פרט אחד בה עשוי להשתנות, ובכך עדיין לא התחלנו לעבוד. אמרנו שניצול הספרות המקצועית דין הוא שייכלל בכרטסת ההבאות לשם נוחיות הכותב. כלומר, שלאחר כל כרטיס ההבאה תבוא הספרות המקצועית הדנה בהבאה זו, כנהוג, למשל, ב־ ThLL. זה לא יוכל להתקיים בשיטה החדשה, שמבנה הכרטיס שונה בה לחלוטין, ועל אחת כמה וכמה לא יוכל להיות, אם יוקלט החומר לתוך גלילים. מצד אחר, אין כל טעם להעביר הכרטיסים שיושגו מן הניצול של הספרות אל השיטה החדשה – זהו בזבוז כספים ללא צורך. לכרטסת של ניצול הספרות יש רק ערך מסייע, שייוודעו, או שלא יישכחו, הדעות והקביעות לגבי ערך מסוים בשעה שייכתב, ולא עוד. מה שאין כן כרטסת ההבאות, שהיא כשלעצמה קניין יקר לכל עבודה בלשנית, וערכה מוחלט. שכן כלולים בה עדי השימוש של הלשון, ואלה אינם משתנים לעולם עם שום שינוי שיטה.

לפיכך תהיה כרטסת הניצול של הספרות אוטונומית, ויהא אפשר להתחיל בה בהקדם, כמובן לאחר שייקבעו כללים ברורים, אילו ספרים ואילו כתבי עת יש לנצל, ומאימתי, וייתן דפוס לדרך הניצול. כל זה מחייב מחשבה מסוימת, ויש לכך חומר בידנו.

אינו באים עכשיו, ואף לא בזמן הקרוב, לכתוב כרטיסים מן המקורות, אבל עבודתנו צריכה להתכוון לשלב זה. בארצנו יש כל אותן המכונות, שמשתמשים בהן בארצות האחרות גם לעבודה הבלשנית. ודאי, בתחילה נהיה מוכרחים לעשות את עבודת השלב השני מחוץ לכותלי האקדמיה, אבל לאורך ימים לא יוכל מוסד לקסיקוגראפי לוותר על אחת או שתיים מן המכונות, נקווה שעד אז יהיו זולות יותר, ויהא בהישג ידה של האקדמיה לקנותן.

שמא לא תהיה זו העזה יתרה להביע בהזדמנות זו את הרעיון, שהאקדמיה ללשון העברית, שנכנסה בעול מפעל כה גדול כשל המילון ההיסטורי, תהיה יוזמת ושותפת להקמת המרכז לניתוח ספרותי ובלשני כדוגמת המרכזים באיטליה בצרפת והמוקמים בגרמניה ובמקומות אחרים. יש שמרכז כזה מוקם בשיתופם של מכונים שונים, מכון למתימטיקה, מכון לפילוסופיה ומכון לבלשנות. מה שלא ניתן לאדם או למוסד לעשות לבדו, אפשר לעתים קרובות לעשותו יחד, ותיטול האקדמיה את חלקה הנוגע ללשון העברית, ותהיה מספקת את הדוקומנטציה הלשונית לכל העולם העברי, כשם שביזאנסון מספקת אותה בתחום הלשון הצרפתית לכל אוניברסיטאות צרפת ומחוצה לה. הרי ממילא מפעל המילון ההיסטורי חייב להתפתח לכלי, שישרת את מחקר הלשון העברית לכל תחומיו ולכל היקפם.

גד בן־עמי צרפתי

המיכון של חקר הלשון

ברשימה זו ניתנים ראשי פרקים על דרכי העבודה ועל המפעלים של שני מוסדות – האחד באיטליה והאחד בצרפת – שהוקמו למטרה המיוחדת של פיתוח המיכון של האנליזה הספרותית – והם באים כהדגמה וכתוספת ביאור למאמרו של פרופ׳ בן־חיים.

החומר המוגש כאן נאסף מחוברות ומעלונים שונים, ובעיקר מן הפרסומים האלה:

1. R. Busa, Summary of the experience of the Centro per l'Automazione dell'Analisi Letteraria of the Aloisianum, Gallarate (Varese), Italia.

(הוגש ל״קולוקביום על השיטות המיכניות של האנליזה הספרותית ושל: הלקסיקוגרפיה״, טיבינגן, 24–26 בנובמבר 1960.)

2. R. Busa, Le fasi dell'Automazione dell'Analisi Letteraria.

3. Spécimens de travaux lexicographiques et lexicologiques réalisés par le Laboratoire d'Analyse Lexicologique, Centre d'Etude du Vocabulaire français, Besançon, 1960.

האמצעים שמשתמשים בהם במיכון המחקר הלשוני הם משני סוגים: א) מכונות חישוב בכרטיסים מנוקבים; ב) מחשבים אלקטרוניים.

הפעולה הראשונה בעבודתן של מכונות החישוב בכרטיסים מנוקבים היא כתיבת הטקסט על ידי המכונה המנקבת, אשר מערכת מקשיה דומה לשל מכונת כתיבה רגילה. בדרך זו הטקסט מועבר שורה שורה לכרטיסים מנוקבים שאפשר לבצע בהם פעולות של הכפלה, מיון וסידור בעזרת מכונות שונות. מערכת המכונות האלה כוללת: מבקרת, מפענחת, ממיינת, מזווגת, מעתיקה, ומלווחת. מכשירים אלו, או דומים לאלו, נמצאים היום בכל החברות המסחריות הגדולות ובמשרדים הציבוריים הגדולים (אף בישראל) לצורך עבודות של סטטיסטיקה ושל חשבונאות, ואפשר בהם להשתמש אף למחקרי לשון וספרות.

יתרון הכרטיסים המנוקבים הוא אפוא, שספירתם, סידורם, מיונם, הכפלתם והדפסתם נעשים אוטומטית – לפי קריטריון הניתן לבחירה ולקביעה על ידי החוקר, מתוך קריטריונים רבים שהכרטיסים מאפשרים. השיטה המיכנית מאפשרת אפוא ניצול של חומר רב ללא חשש של שכחה וטעות, ארגון מהיר של החומר הנאסף, מיזוג של סדרות כרטיסים שונות, או הפרדה של כרטיסים המכוונים לדרישה מסוימת, הפצת החומר (ע״י שכפול או הדפסה) – וכל זה במחיר שאינו עולה על מחירן של השיטות המקובלות.

בדרך זו המכונה מקטינה במידה עצומה את הזמן, שהחוקר צריך להקדיש ל״עבודה שחורה״ של חיפוש, רישום, סידור, חישוב, והיא דורשת ממנו ניצול מתמיד ונמרץ יותר של סגולותיו לניתוח, להבנה ולארגון.

לא נעסוק כאן בתיאור עבודת המחשבים האלקטרוניים: למעשה השגיהם שווים לשל המכשירים מן הסוג המתואר לעיל. נזכיר רק לדוגמה, כי באוניברסיטה של מנצ׳סטר נערכה על ידי מחשב אלקטרוני רשימת מלים של הטרגדיה Mithridate לראסין. הרשימה כוללת את כל המלים שבטרגדיה לצורותיהן השונות בסדר האלף־בית; על יד צורה החוזרת בטקסט לא יותר מ־64 פעמים, בא ברשימה זו מראה מקום שלם, ועל יד צורה שכיחה יותר בא רק המספר המציין את תדירותה.

א. המרכז למיכון האנליזה הספרותית בגאלאראטי (איטליה)

המרכז למיכון של האנליזה הספרותית נוסד בגאלאראטי, בקרבת מילאנו, למען הגשמת הרעיון של Index Thomisticus, הוא קונקורדנציה לכתבי תומס מאקווינו, הכוללים למעלה מעשרה מיליון מלים. הניסיונות המעשיים הראשונים נעשו בשנת 1949, ומאז עד עתה עסק המרכז בעריכת קונקורדנציות ורשימות מלים לכמה חיבורים לפי קריטריונים שונים והביא אותם לשלבי ביצוע שונים. עד סוף אוגוסט 1960 עובדו 20 טקסטים ספרותיים, ובהם הוולגאטה, מגילות ים המלח, Farben-lehre של Goethe, מלבד כמה מחיבורי תומס מאקווינו – בסך הכול כמעט שלושה מיליון מלים בשלושה כתבים (לאטיני, יווני, עברי) ובשבע לשונות (יוונית, לאטינית, איטלקית, גרמנית, עברית, ארמית, נבטית).

אלו הם שלבי העבודה במוסד זה:

א. עריכה מוקדמת

החוקר עורך את הטקסט, ומוסיף בו את ההוראות המתאימות לניקוב.

ב. כרטיסי טקסט (text-cards)

א) כתבנית מעתיקה את הטקסט במכונה המנקבת, ומייצרת על ידי זה את כרטיסי הטקסט. בכרטיסים אלו באים כל המשפטים (או כל השורות) של הטקסט, משפט אחד (או שורה אחת) על כל כרטיס, כתובים על ידי נקבים לפי מפתח אלפביתי מוסכם; בראש כל כרטיס בא מראה מקום המשפט בטקסט, אחרי כן מספר סידורי, וציון מיוחד (עליו להלן).

ב) כתבנית אחרת בודקת באופן מיכני את כרטיסי הטקסט שנכתבו ומחליפה את המשובשים.

ג. כרטיסי מלים (word-cards)

מכרטיסי הטקסט מייצרת המכונה באופן אוטומטי את כרטיסי המלים, שכל אחד מהם נושא מלה אחת. מכאן שמתוך כרטיס טקסט המכיל משפט אחד מייצרת המכונה כרטיסי מלים כמספר המלים שבאותו משפט; בסך הכול יהיה מספר כרטיסי המלים כמספר המלים של החיבור.

בהמשך פעולה זו מצרפים לכל מלה באופן אוטומטי את הציונים הבאים (או ציונים אחרים ואף רבים יותר, לפי תכנית העבודה):

1. מראה מקום המלה.
2. האות הראשונה של המלה שלפניה והאות הראשונה של המלה שלאחריה.
3. המספר הסידורי, המציין את מקום המלה בטקסט.
4. הציון המיוחד של המשפט שהמלה שייכת אליו (אם יש למשפט ציון מיוחד).

ציונים מיוחדים אלה נקבעים מראש, כדי לציין שהמשפט הוא מובאה מדברי מחבר אחר או מובאה מחיבור אחר של אותו מחבר וכיצא בהם.

על גב הכרטיס ברווחים האופקיים שבין שורות הנקבים אפשר להדפיס באותיות רגילות את הקטע שאליו שייכת המלה (עד 12 שורות, שכל אחת מהן מכילה 64 סימנים).

ד. כרטיסי צורות (form-cards)

מתוך כרטיסי המלים מייצרת המכונה אוטומטית, ללא התערבות של עובד, את כרטיסי הצורות. הם מכילים את כל הצורות השונות הנמצאות בחיבור, כל אחת רק פעם אחת. (דוגמה לביאור: אם בחיבור נמצאות שש המלים, "דבר", "דבר", "דברי", "דברים", "דברים", "דברים", הן יתנו שישה כרטיסי מלים, ומהם יתקבלו שלושה כרטיסי צורות: "דבר", "דברי", "דברים"). בשלב עבודה זה מחשבים כ"מלה" קבוצת סימנים רצופים, שיש רווח לפניה ואחריה. תופסים כאן את המלה רק לפי אופייה הגראפי. המלים "הלכתי", "אלך", "לכת", למשל, יופיעו כאן כשלוש מלים שונות: הראשונה תסודר עם המלים הפותחות באות ה, השנייה בין הפותחות בא׳, השלישית בל׳. לעומת זה קבוצת האותיות "שמר" תיחשב כמלה אחת, אף על פי שהיא עשויה להיות עבר נסתר, ציווי, מקור, ועוד. אין המכונה מבחינה בין ההומוגראפים.

כמו כן "בבית" תיחשב כמלה אחת, ואילו "היה אומר" יחולק לשתי מלים.

בכל כרטיס מן הכרטיסים האלה מצטרפים אוטומטית למלה ציון תדירותה והמספר המציין את המקום שהיא תוספת בסדרן של כל המלים של הטקסט המסודרות לפי האלף־בית.

לתוצאה זו מגיעים על ידי התהליך הזה:

א) המכונה מסדרת את כל כרטיסי המלים בסדר האלף־בית;

ב) המכונה בודקת את כל כרטיסי המלים במהירות של 6,000 לשעה והיא מדפיסה על דף רק את המלים השונות שהיא פוגשת.

כך היא מלאכתה: היא מדפיסה את המלה הראשונה; אם המלה שלאחריה שונה ממנה, הריהי מדפיסה גם אותה; אם היא שווה לה, אין המכונה מדפיסה אותה אלא סופרת אותה; לאחר שספרה את כל המלים השוות, היא מדפיסה את המספר הכולל על יד המלה המודפסת כבר. אחר כך היא מדפיסה את המלה השונה הבאה מיד אחר כך, וכך היא ממשיכה עד הסוף. כשהמכונה עורכת בדפוס רשימה מסכמת זו של אוצר המלים של מחבר מסוים, באותה שעה היא אף מנקבת כרטיסים נפרדים לכל מלה שונה ורושמת על כל כרטיס את תדירות המלה ואת מספרה הסידורי בסדר האלף־בית (מספר סידורי ראשון).

ה. כרטיסי ערכים (main-cards)

א) החוקר בודק את הרשימה של כרטיסי הצורות ומקבץ את כל הצורות השונות של יחידה גראפית־סימאנטית אחת, לערך אחד במילון. הוא יקבץ, למשל, הלכתי, אלך, לכת בערך "הלך".

ב) יתר על כן, החוקר מבדיל את ההומוגראפים לפי הוראותיהם השונות. "ערב" למשל, יכול להיות: שקעה החמה, קיבל על עצמו אחריות, נָעַם (הוראות אלו –

לפי הניקוד עָרַב; אולם אם לא תרשום המכונה את הניקוד, מספר ההוראות יהיה גדול הרבה יותר).

ג) החוקר יערוך את רשימת הערכים.

ד) ערכים אלו רושמים אותם אחד אחד על כרטיסים, מסדרים אותם לפי סדר האלף־בית, ומסמנים אותם במספרים לפי סדרם, ועל ידי זה מקבלים את כרטיסי היסוד או כרטיסי הערכים.

הם מייצגים רשימה מסכמת שנייה של אוצר המלים של המחבר המסוים, לא רק על יסוד גראפי של המלים, אלא על יסוד גראפי־סימאנטי.

ה) אחר כך מצרפים את כרטיסי הצורות ואת כרטיסי המלים לכרטיסי הערכים המתאימים. עם זה יירשם אוטומטית המספר הסידורי השני בכל כרטיסי המלים ובכל כרטיסי הצורות.

בו בזמן תוסיף המכונה בכרטיסי הערכים את ציון התדירות הכוללת, היא מספר הפעמים, שערך מופיע בחיבור לצורותיו השונות.

אם לא נעשה הדבר מקודם, "מפענחים" עתה את ארבע קבוצות הכרטיסים: המכונה תדפיס על כל כרטיס, לאורך צדו העליון, באותיות וספרות רגילות, מה שרשום בו על ידי נקבים.

על ידי זה יצרה המכונה, על יסוד הניקוב והאימות שבתחילה, שני העתקים שלמים של הטקסט (כרטיסי טקסט וכרטיסי מלים) ושתי רשימות מסכמות של אוצר המלים של המחבר (כרטיסי צורות וכרטיסי ערכים).

ו. רשימת מלים (index verborum) וקונקורדנציה

הכרטסת שנתקבלה בדרך זו תוכל ראשית כול להדפיס, על גבי דפים או בצורת קונטרסים או כרכים, "רשימות מלים" ו"קונקורדנציה" – כלי עזר חשובים למחקר פילולוגי.

לשם זה דיינו, אם נכניס לתוך המכונה את כרטיסי המלים יחד עם כרטיסי הערכים או עם כרטיסי הצורות או עם שתי המערכות כאחת, והיא תדפיס בצורה מוגמרת, בעמודים מסודרים ובריווחים נכונים, ללא צורך בהגהה, את כל העמודים, אשר יהוו את הקונקורדנציה של החיבור. אפשר להכפיל אחר כך את העמודים בליתוגראפיה או בדרך אחרת.

מהירות ההדפסה היא מ־4,800 עד 9,000 ועד 60,000 שורות לשעה, לפי סוג המכונה שמשתמשים בה.

ב. המעבדה לאנליזה לקסיקולוגית שעל יד המרכז המחקרי של המילון הצרפתי בביזאנסון (צרפת)

המעבדה לאנליזה לקסיקולוגית נוסדה בשנת 1959 כחלק של המרכז המחקרי של המילון הצרפתי שעל יד האוניברסיטה של ביזאנסון. מגמת המעבדה היא אוסף העדויות המילוניות הקיימות והשלמתן, תיאום עבודות המחקר השונות בתחום זה, ויצירת "כרטסת כללית של המילון הצרפתי". המעבדה שואפת לשיתוף פעולה עם מוסדות מחקר אחרים, הן בצרפת הן מחוצה לה. היא תוציא לאור במונוגראפיות או במאמרים שבתוך מאספים את המחקרים שייעשו בה או בתיאום עמה והדנים באוצר המלים של הלשון הצרפתית ובתולדותיו (תעודות לקסיקוגראפיות, רשימות מלים

של טקסטים ספרותיים, של מחברים, של מקצועות ושל תקופות) – ומלבדם היא מפרסמת כתב עת Les Cahiers de Lexicologie וגם Bulletin d'Information. החומר שלא יתפרסם בדפוס, ובייחוד הכרטסות, יוכפל ויימסר למרכזים ולמוסדות דומים לשימוש החוקרים.

האמצעי הטכני המיוחד של המעבדה היא הכרטסת המיכנית, בדומה למרכז של גאלאראטי, אלא שכאן העבודה מתחילה בניקוב סרט, ולא בניקוב כרטיסים.

מכונות החישוב וההדפסה הרגילות הותאמו במיוחד לעבודה ספרותית, באופן שמספר הסימנים האפשריים הוגדל מ־36 עד 59, והם מספיקים לאותיות האלף־בית הלאטיני בתוספת האקצנטים השונים הנהוגים בלשון הצרפתית, לסימני הפיסוק ולמספרים. הציוד הטכני של המעבדה כולל: שלוש מנקבות סרטים – קוראת סרטים אחת – שתי קוראות־מנקבות – מפענחת אחת – מפענחת־מעבירה אחת – מעתיקה אחת – מערכת אחת לקריאה פוטו־אלקטרית – ממיינת אחת – מזווגת אחת – מלווחת אחת – שתי מדפיסות (59 סימנים).

במעבדה נעשות עתה הפעולות הבאות:

א. רשימת מלים

מגמתה היא עריכת מפתח שיטתי של כל המלים של טקסט מסוים בצירוף מראי המקומות של כל מלה בטקסט. שלבי העבודה הם:

1. ניקוב הטקסט – נעשה על ידי כתבנית, על סרט;

2. עריכת כרטסת – מתוך הסרט המנוקב נעשים באופן מיכני כרטיסי מלים; כל אחד נושא מלה אחת, רשומה בנקבים ואף כתובה באותיות רגילות, בצירוף מראה מקום המלה בטקסט; הכרטיסים נסדרים אוטומטית בסדר אלף־בית גרידא, ונסדרים שנית בסדר אלף־בית, אחר מיון ההומוגראפים והצורות הדקדוקיות השונות של כל מלה;

3. הדפסה של רשימת המלים – נעשית באופן אוטומטי, בצירוף המספר הסידורי ומספר המורה על תדירותה של המלה. בטקסטים פיוטיים אפשר לערוך גם רשימות המלים החורזות.

ב. אוספים לקסיקולוגיים

כרטסת מילונית עשויה להיות בסיס למחקרים בלשניים, היסטוריים וספרותיים שונים. אחרי ההשקעה הראשונית ביצירת הכרטסת, ידרוש כל מחקר רק השקעה מועטת מאחר שאין צורך בכתיבת כרטסת חדשה לפי כל קריטריון חדש של מחקר, שכן אותה כרטסת ניתנת לסידור מיכני חדש במהירות רבה ואף להכפלה מיכנית.

כרטיסי האוסף הזה נושאים:

1. ציון מספרי של הכרטיס
2. מראה מקום המלה
3. המלה
4. פרטים דקדוקיים:
איזה חלק מחלקי הדיבור, מין ומספר, צורה דקדוקית, מלים המלוות את המלה, לגבי פועל – עומד או יוצא
5. ידיעות סימאנטיות:
התחום הסימאנטי של המלה: מלה של מקצוע מסוים, של שכבה לשונית מיוחדת

6. תאריך:
תאריך הוצאת הספר שממנה לקוחה המלה

7. פרטים על המחבר:
תקופתו, שכבתו התרבותית, השתייכותו החברתית, מוצאו הגיאוגרפי, דתו וכו׳

8. פרטים על החיבור:
פרוזה, פרוזה אמנותית, פרוזה שירית, שירה חופשית, שירה, טסקט ספרותי, טקסט לא־ספרותי או טכני, המקצוע שהטקסט עוסק בו, וכו׳

9. פרטים על הקטע:
הבאה, סיפור, דיבור ישיר או בלתי ישיר, תשובה; מעמדו החברתי של המדבר, גילו, משלח ידו; נוסח אחר או הגהה

10. פרטים על המלה:
כתיבה באותיות מיוחדות, באותיות גדולות, בין מרכאות, וכו׳; כתיב בלתי רגיל, צורה משנית, מקום המלה במשפט או בחרוז; מלה חורזת, מלת פתיחה, וכו׳

11. מספר סידורי של המלה.

הארגון השונה של הידיעות שעל גבי כל כרטיס, והידיעות שאפשר להוסיף על ידי כרטיסים נוספים, ניתנים לשיקול המומחים.

גם כרטסת זו נעשית בשלבים: רישום הטקסט על סרט מנוקב – עריכה מיכנית של כרטיסי טקסט – עריכה מיכנית של כרטיסי מלים. ראוי לציין, כי באמצעים טכניים פשוטים ניתן להגביל את כרטסת המלים ל״מלים מורות״ בלבד (מלים בעלות ערך סימאנטי, כגון רוב השמות והפעלים) או, להפך, ל״מלים־כלים״ בלבד (מלות קישור, מלות יחס וכיוצא בהן), ובזה משיגים חיסכון, בלי להשמיט כל חומר שהחוקר מעוניין בו.

הניצול המילוני של הכרטסת הוא רב־גוני ביותר; ממנה אפשר לקבל אוטומטית, למשל, כרטסת של ״לשון הרומאן במחצית הראשונה של המאה הי״ט״, או כרטסת של ״שמות הצבעים במילון של האופנה הפאריזאית בין 1830 ל־1832״.

ג. אוצר המילונים הצרפתים

ניצול המילונים הצרפתים בני התקופות השונות ויצירת כרטסת מתאימה, אשר תסכם, לפי סימנים מוסכמים, את הידיעות השונות, שהמילונים מוסרים על כל מלה. בעזרתה אפשר יהיה לערוך מחקרים על מעברי מלים משכבה לשונית אחת לשכבה אחרת, מסגנון לסגנון, מניב לניב.

ד. מפתח היסטורי לאוצר המלים של הצרפתית

המילונים השונים מציגים הבאות שונות למלים וקובעים תאריכים להופעה הראשונה של כל מלה בלשון הכתובה. למילונים מצטרפים גם מחקרים שונים כמקורות של ידיעות כרונולוגיות על המלים. אפשר לבנות כרטסת, אשר תרכז את התאריכים השונים, הניתנים במקורות לכל מלה ומלה, ;תאפשר על ידי כך את קביעת התאריך המוקדם ביותר של כל מלה ואף את תיקון התאריך לפי המחקרים המתפרסמים מדי פעם בפעם. אפשר יהיה לשלב למילונים גם אוספים של חוקרים אשר עדיין לא נדפסו ולא נוצלו.